Chapter 1
Detecting Abnormal Traffic in Wireless Networks Using Unsupervised Models

Alexis Huet
Nanjing Howso Technology, China

ABSTRACT
Development of high-speed LTE connections has induced an increasingly quantity of traffic data over the network. Detection of abnormal traffic from this continuous stream of data is crucial to identify technical problem or fraudulent intrusion. Unsupervised learning methods can automatically describe structure of the data and deduce patterns of the network. There are useful to identify unexpected behaviors and to promptly detect new type of anomalies. In this article, traffic in wireless network is analyzed through different unsupervised models. Emphasis is given on models combining traffic data with time stamps information. A model called Gaussian Probabilistic Latent Semantic Analysis (GPLSA) is introduced and compared with other methods such as time-dependent Gaussian Mixture Models (time-GMM). Efficiency and robustness of those models are compared, using both sampled and LTE traffic data. Those experimental results suggest that GPLSA can provide a robust and early detection of anomalies, in a fully automatic, data-driven solution.

DOI: 10.4018/978-1-5225-1750-4.ch001

Copyright ©2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

Data monitored through telecommunication networks have grown exponentially in the past few years. The resulting flow has become impossible to manually process and analyze. In particular, detection of unexpected traffic behaviors from normal patterns has remain an important issue. This field is critical because anomalies can cause deficiency in network efficiency. Indeed, origin of those anomalies can be a technical problem of a cell or a fraudulent intrusion in the network. There are typically urgent to identify and to fix. Consequently, data-driven systems have been developed to identify anomalies, using machine learning algorithms. The purpose is to automatically extract information from raw data, to identify and alert when an anomaly occurs.

In wireless networks, collected data contain values for different features as well as time stamps. To seek and detect anomalies, values can be modeled and processed using unsupervised algorithms. This kind of algorithms assumes that information about which elements are anomalies is unknown, by using unlabelled data. This is usually the case, since anomalies in the traffic data are rare and may take many forms. Unsupervised algorithms automatically separate and distinguish data structures and patterns. They do not intend to directly detect anomalies, but only to describe and group data. Afterwards, zones of anomalies are deduced from those groups. The main advantage of this methodology is the ability to detect previously unseen or unexpected anomalies.

Another component to take into consideration for wireless networks data is time stamps. This information is commonly collected when data are generated but is not widely used in classic anomaly detection processes. However, network load has daily fluctuations. For example, a normal value during peak period may be an anomaly outside, and remains undetected. Adding time stamp attributes in a model allows to discover periodic behaviors.

In this article, unsupervised models are used to detect anomalies. Specifically, the following sections focus on algorithms combining both values and dates. For this purpose, two models are introduced. The first one is time-GMM, which is a time-dependent extension of GMM (McLachlan & Basford, 1988) by considering each period of time independently. The second one is GPLSA (Hofmann, 1999b), which combines together values and dates processing in a unique machine learning algorithm. This latter algorithm is well-known in text-mining and recommender systems areas, but was barely used in other domains such as anomaly detection. Those algorithms are implemented and tested on sample and traffic data. Their ability to find anomalies and to adapt to new patterns is shown. Robustness, complexity and efficiency of the algorithms are compared.
Mobility and Enterprise 2.0
[www.igi-global.com/article/mobility-enterprise/37545?camid=4v1a](www.igi-global.com/article/mobility-enterprise/37545?camid=4v1a)

Cross-Layer Design for Packet Data Transmission in Maximum Ratio Transmission Systems with Imperfect CSI and Co-Channel Interference
[www.igi-global.com/chapter/cross-layer-design-packet-data/65678?camid=4v1a](www.igi-global.com/chapter/cross-layer-design-packet-data/65678?camid=4v1a)